

# Análise de séries temporais

por similaridade e alinhamento não  
linear com Dynamic Time Warping



Tarefas de mineração de  
séries temporais por  
similaridade

# Similaridade

# Similaridade em séries temporais

"**Similarity** may not be the best approach. But, **for every task in time series mining**, there is at least one very simple **similarity-based algorithm that performs better than 95% of other methods.**"

Eamonn Keogh

UCR



# Similaridade

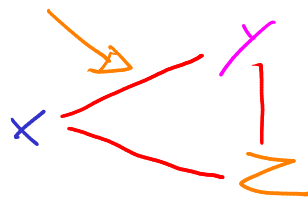


# Similaridade vs distância

São coisas contrárias, mas é comum falarmos “similarity-based” e usarmos algoritmos que se baseiam em **distância**.

Mas vou continuar me referindo aos algoritmos como baseados em similaridade às vezes.

# Similaridade



Propriedades interessantes (porém não obrigatórias) de medidas de distância

- $d(x,y) \geq 0$  não-negatividade
- $d(x,y) = 0 \Leftrightarrow x = y$  identidade dos indiscerníveis
- $d(x,y) = d(y,x)$  simetria
- $d(x,y) \leq d(x,z) + d(z,y)$  desigualdade triangular ← Indexar

Tendo essas propriedades, podemos dizer que essas funções são métricas.

Exs: **distância euclidiana** (DE), distância do cosseno, etc

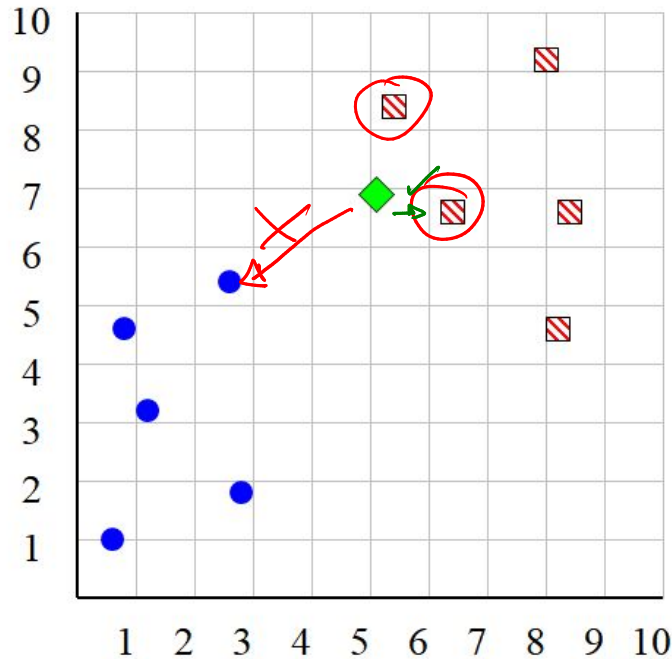
$$\sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

~> tira a raiz

$$\sqrt{A} > \sqrt{B}$$
$$A > B$$

# Aprendizado baseado similaridade - NN

Nearest Neighbor



5

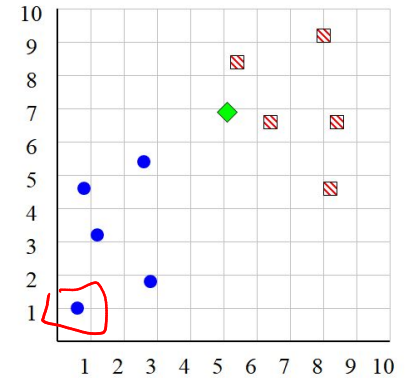
5

? →



# Aprendizado baseado similaridade - NN

id	a1	a2	classe
1	0.5	1	1
2	2.9	1.9	1
3	1.2	3.1	1
4	0.8	4.7	1
5	2.7	5.4	1
6	8.1	4.7	2
7	8.3	6.6	2
8	6.3	6.7	2
9	8	9.1	2
10	5.4	8.4	2
11	5	7	?



Retirado do tutorial: A Gentle Introduction to Machine Learning and Data Mining to the Database Community - Eamonn Keogh, SBBD 2003

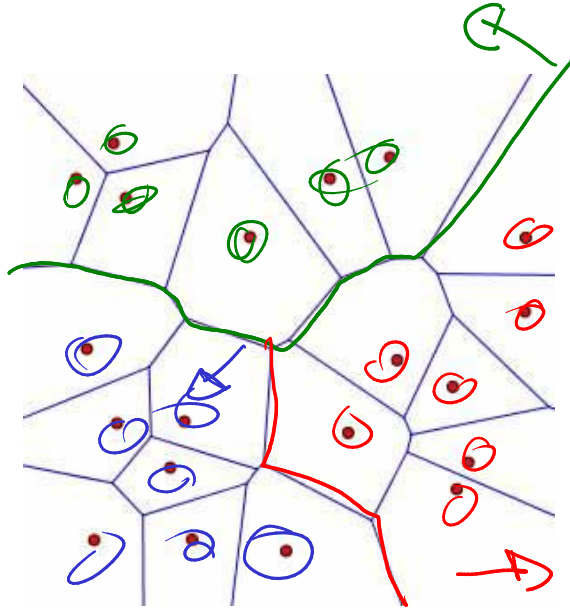
# Aprendizado baseado similaridade - NN

<del>11</del>	a1	a2	classe	DE (11,i)
1	0.5	1	1	7.5
2	2.9	1.9	1	5.5
3	1.2	3.1	1	5.4
4	0.8	4.7	1	4.7
5	2.7	5.4	1	2.8
6	8.1	4.7	2	3.8
7	8.3	6.6	2	3.3
8	6.3	6.7	2	1.3
9	8	9.1	2	3.6
10	5.4	8.4	2	1.4
11	5	7	2	

# Aprendizado baseado similaridade - NN

É possível desenhar as fronteiras de decisão

- Diagrama de Voronoi



# Aprendizado baseado similaridade - kNN

Vamos introduzir o hiperparâmetro  $k$

- Em vez de escolher apenas 1 vizinho, olharemos para  $k$
- $k$  é definido pelo usuário, embora possa ser estimado a partir dos dados



Esse hiperparâmetro (assim como qualquer outro) pode ter grande impacto no aprendizado

# Aprendizado baseado similaridade - kNN

id	a1	a2	classe	DE (11,i)
1	0.5	1	1	7.5
2	2.9	1.9	1	5.5
3	1.2	3.1	1	5.4
4	0.8	4.7	1	4.7
5	2.7	5.4	1	2.8
6	8.1	4.7	2	3.8
7	8.3	6.6	2	3.3
8	6.3	6.7	2	1.3
9	8	9.1	2	3.6
10	5.4	8.4	2	1.4
11	5	7	2	

Para  $k = 1$



# Aprendizado baseado similaridade - kNN

id	a1	a2	classe	DE (11,i)
1	0.5	1	1	7.5
2	2.9	1.9	1	5.5
3	1.2	3.1	1	5.4
4	0.8	4.7	1	4.7
5	2.7	5.4	1	2.8
6	8.1	4.7	2	3.8
7	8.3	6.6	2	3.3
8	6.3	6.7	2	1.3
9	8	9.1	2	3.6
10	5.4	8.4	2	1.4
11	5	7	2	

Para  $k \neq 2^*$   
E se...?

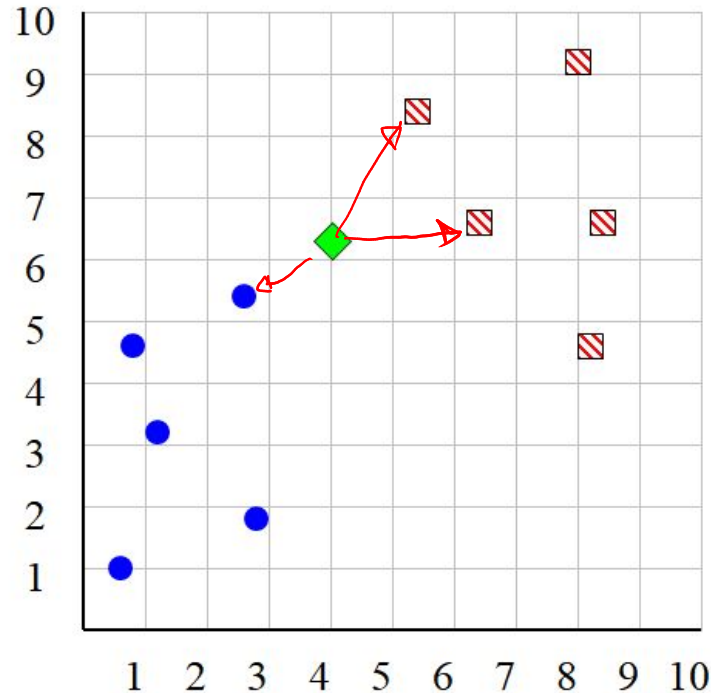
# Aprendizado baseado similaridade - kNN

id	a1	a2	classe	DE (11,i)
1	0.5	1	1	7.5
2	2.9	1.9	1	5.5
3	1.2	3.1	1	5.4
4	0.8	4.7	1	4.7
5	2.7	5.4	1	2.8
6	8.1	4.7	2	3.8
7	8.3	6.6	2	3.3
8	6.3	6.7	2	1.3
9	8	9.1	2	3.6
10	5.4	8.4	2	1.4
11	5	7	2	

Para  $k = 3$

# Aprendizado baseado similaridade - kNN

E agora?



1NN - azul  
3NN - vermelha



# Aprendizado baseado similaridade - kNN

Alternativas aos “problemas”:

- Ponderar o voto pelo inverso da distância (ao quadrado)
- Ponderar os atributos, de acordo com sua contribuição
- Utilizar diferentes medidas de distância
- Normalizar/padronizar os exemplos
- Indexar (se/quando possível)
- Classificar por range

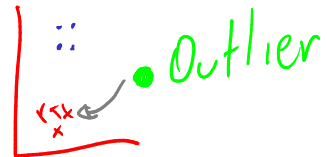
$$\frac{1}{d} = \text{voto}$$

z-norm

$$\hookrightarrow x' = \frac{x - \mu}{\sigma}$$

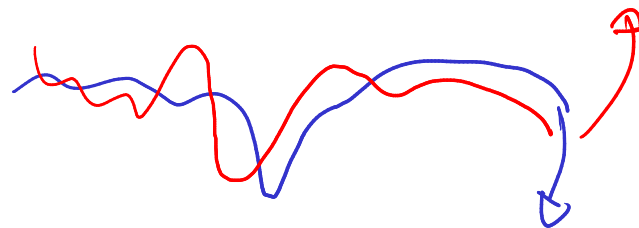
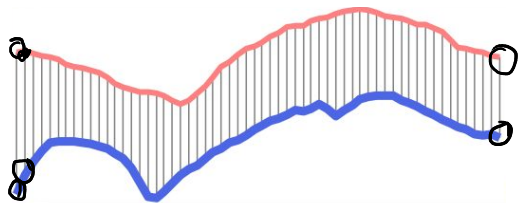
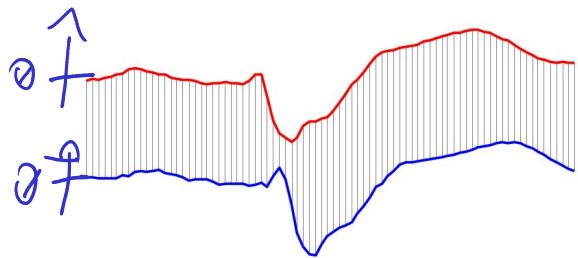
$$\sum \underline{w_i} (p_i - q_i)^2$$

$r \approx 0$



E séries temporais?

# Distância Euclidiana



Alinhamento **linear** entre séries temporais  $x$  e  $y$  de mesmo comprimento  $n$

$$ed(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2}$$

# Exemplos de tarefas

Classificação

Agrupamento

Busca por similaridade

Deteção de anomalia

Descoberta de motifs

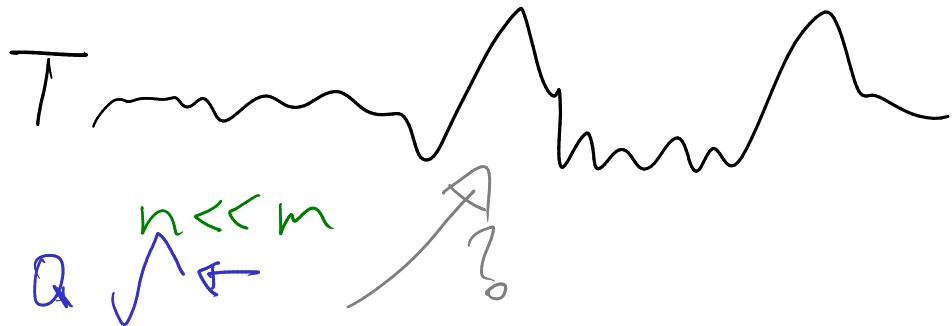
Segmentação semântica

Multidimensional  
Scaling  
MDS

A handwritten diagram illustrating the relationship between 'Multidimensional Scaling' and 'MDS'. The words 'Multidimensional' and 'Scaling' are written on the top line, with an arrow pointing from 'Scaling' to the right. Below this, a horizontal line is drawn, and the letters 'MDS' are written underneath it. To the right of 'MDS', there is a hand-drawn box with a diagonal line through it, and an arrow points from the 'Scaling' part of the top line towards this box.

# Busca por similaridade

- $T$  é uma (longa) série de referência  $m$
- $Q$  é uma série (curta) de consulta  $n$



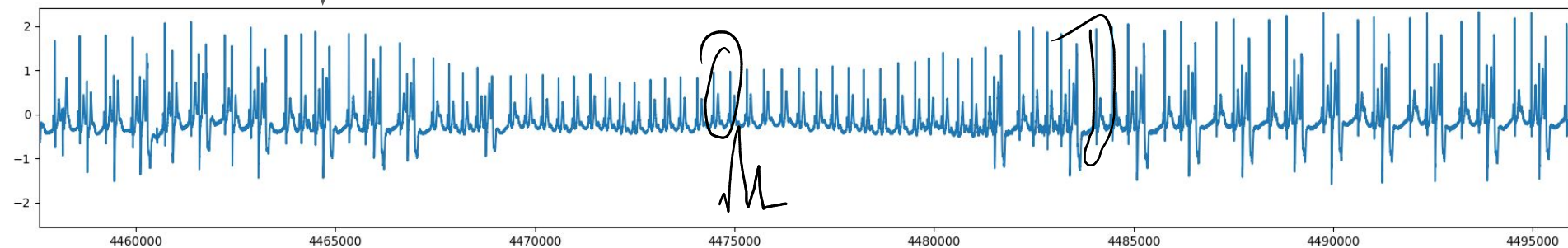
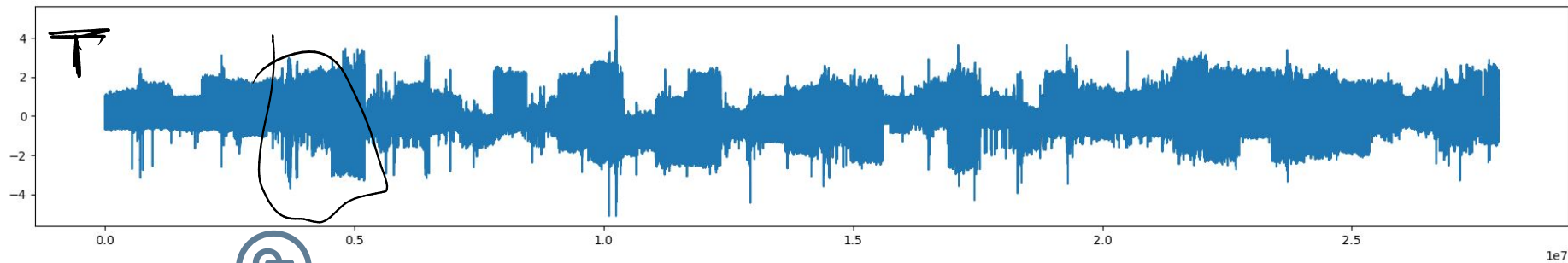
A busca por similaridade consiste em encontrar a subsequência (subconjunto contíguo de observações) de  $T$  que mais se assemelha à série de consulta.

Exemplo: um médico possui um “template” de um batimento cardíaco anômalo que descreve uma determinada doença. Esse batimento ocorre no ECG de um paciente? Em que momento?

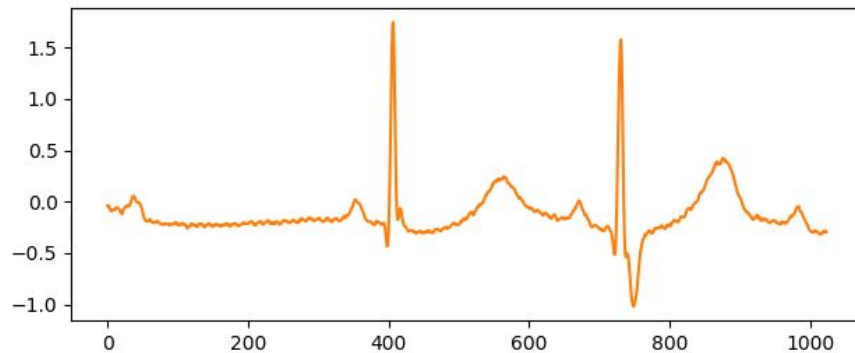
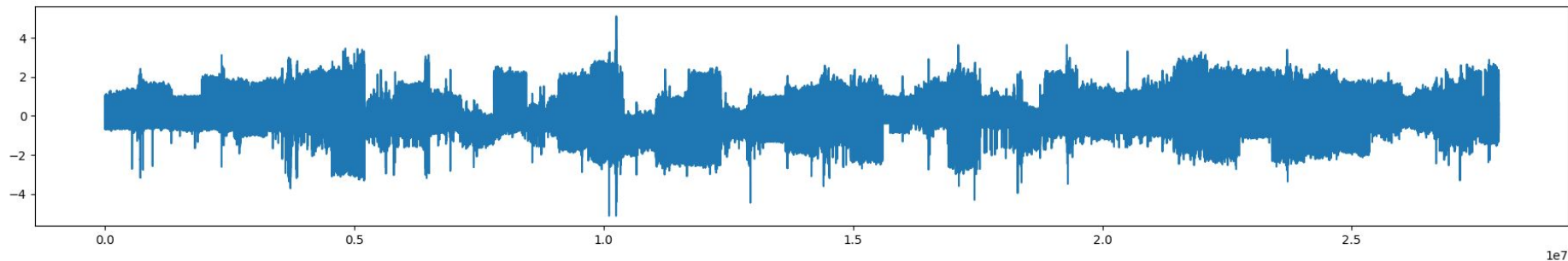
- NN  
↑

- Range  $(d(Q, SS) \leq r)$   
↑

# Busca por similaridade



# Busca por similaridade



? Arritmia

# Busca por similaridade

Prática!



# Busca por similaridade

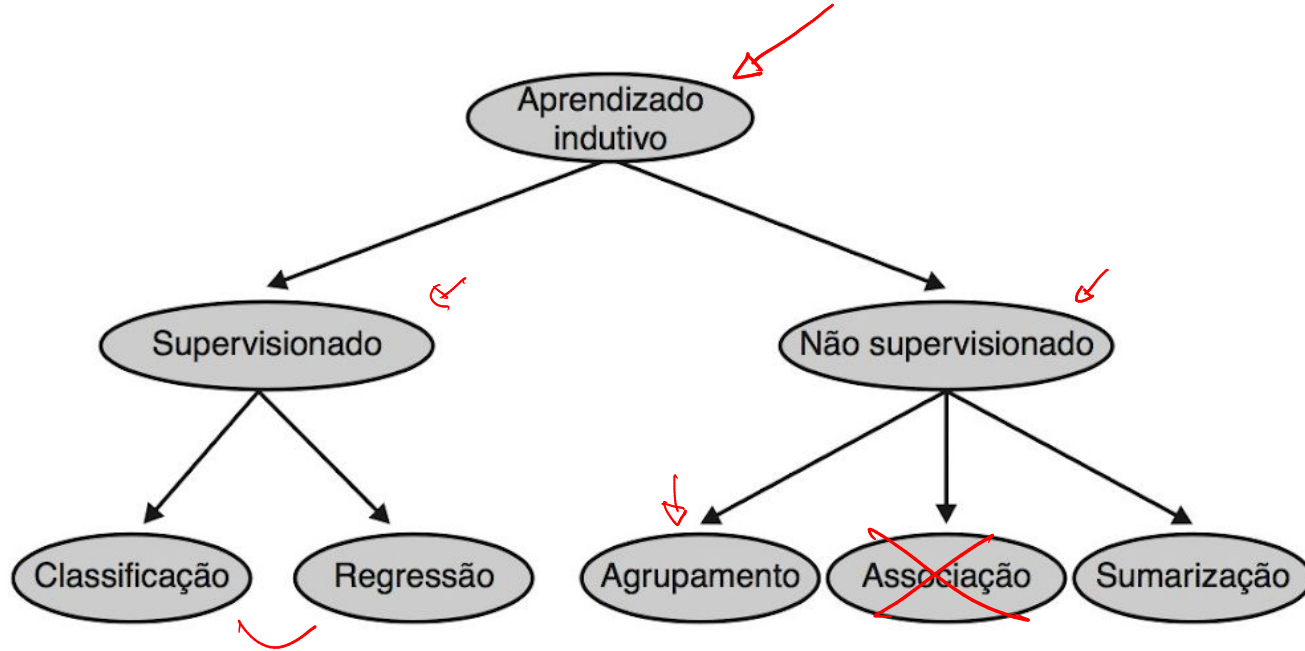
Como isso se estende para outras tarefas?

- Forecasting *prática*
- Classificação
- Agrupamento

# Forecasting

Prática!

# Agrupamento de dados



# Agrupamento de dados

É um assunto que merece um curso inteiro

- Útil para a análise exploratória de dados
- Pode ser uma etapa importante para a tomada de decisão
- Uma das tarefas mais aplicadas em diversas áreas de conhecimento  
(onde é comum ter dados não rotulados)

# Agrupamento de dados

O objetivo da tarefa é encontrar  $k$  ( $k < n$ ) grupos de exemplos (**clusters**) que representem de alguma forma o mecanismo que gerou tais dados. Ou seja, se dois exemplos **representam o mesmo fenômeno ou mesma "categoria"**, eles estão diretamente relacionados. Portanto, devem estar no mesmo grupo.

Mas isso varia e há diversas definições do que é um grupo.

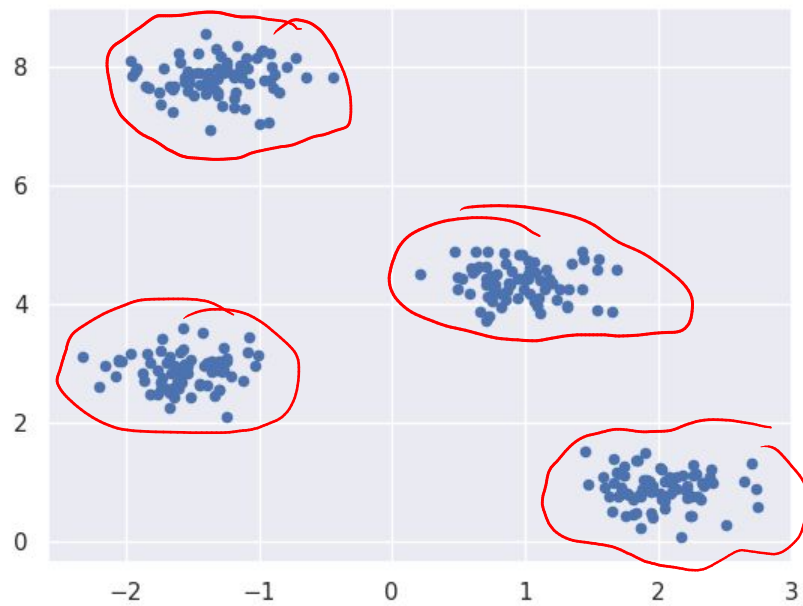
# Agrupamento de dados

Definição bastante comum:

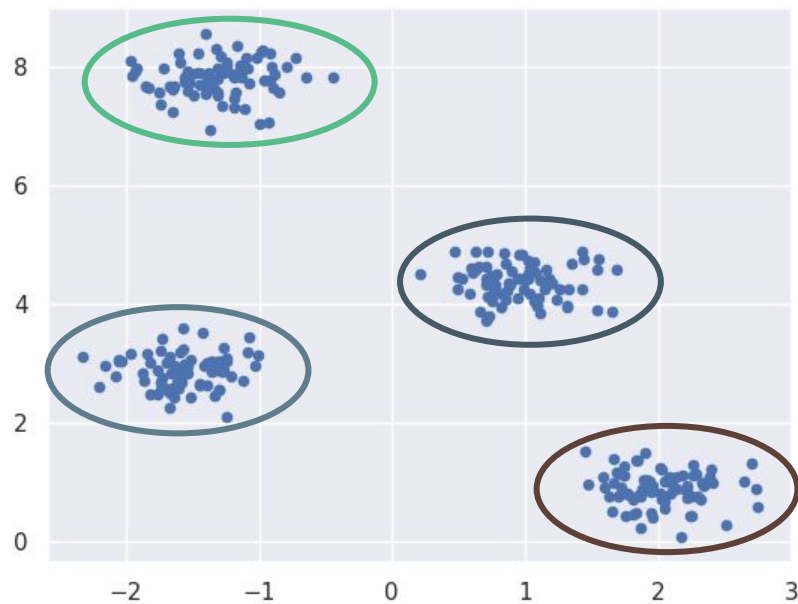
Dois exemplos estão em um mesmo grupo se são similares entre si e distantes de exemplos que estão em outros grupos.

Ou seja, eu pareço com meus vizinhos, mas sou diferente de quem está em outro grupo.

# Agrupamento de dados - exemplos

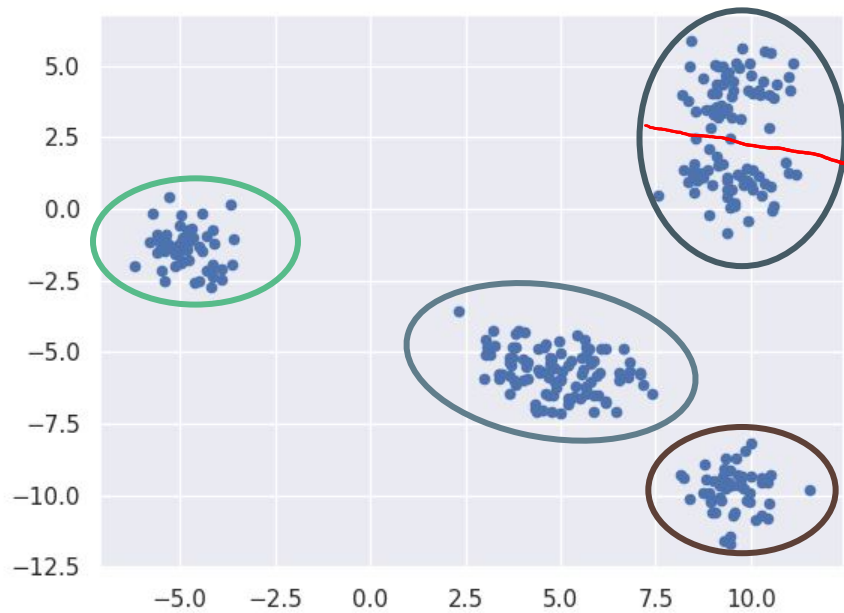


# Agrupamento de dados - exemplos

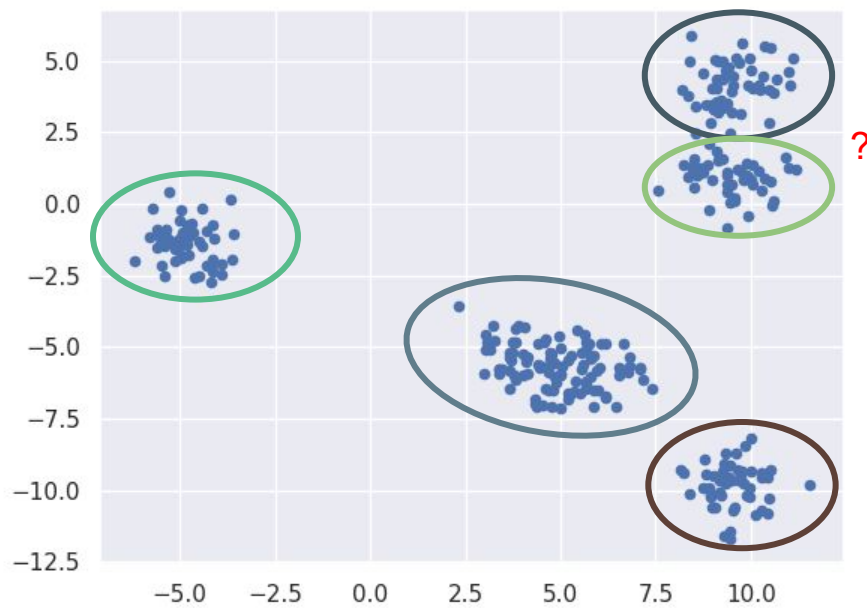




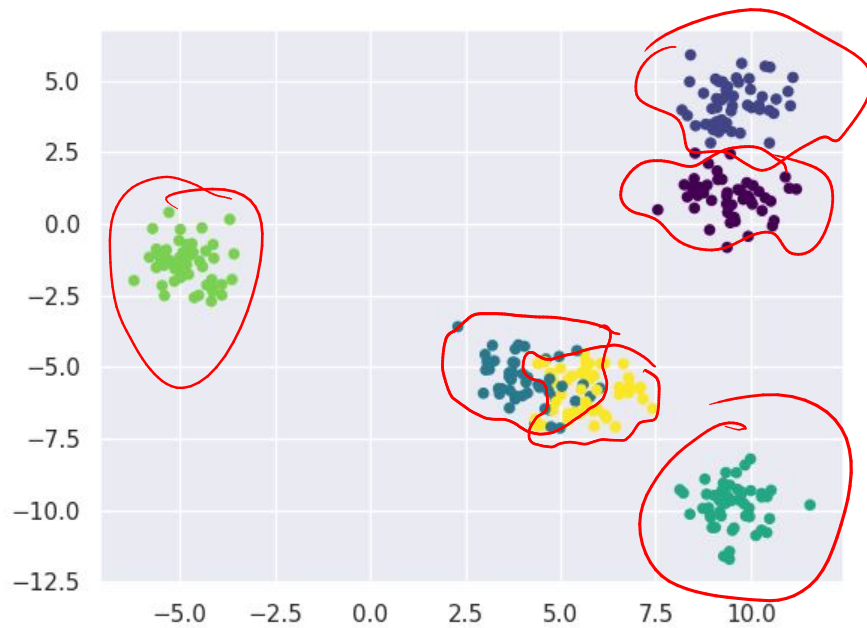
# Agrupamento de dados - exemplos



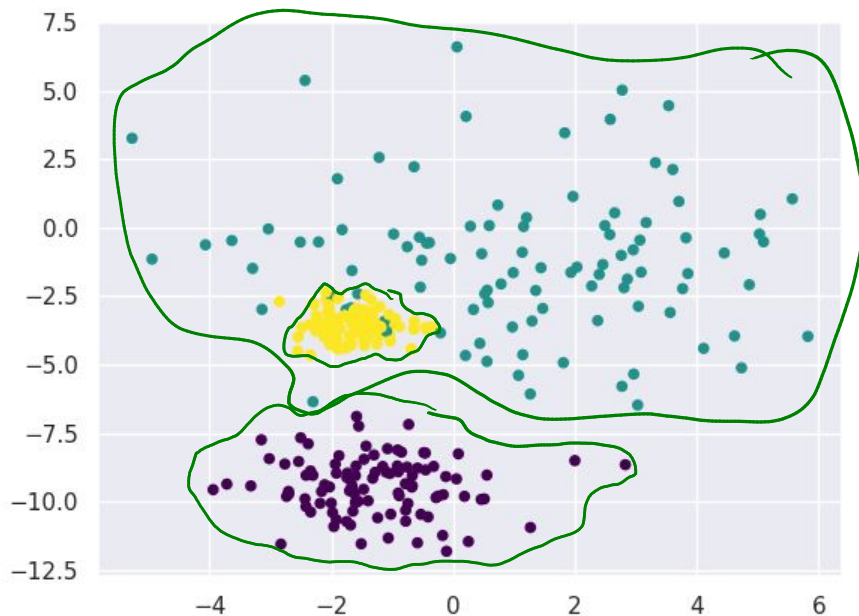
# Agrupamento de dados - exemplos



# Agrupamento de dados - exemplos



# Agrupamento de dados - exemplos

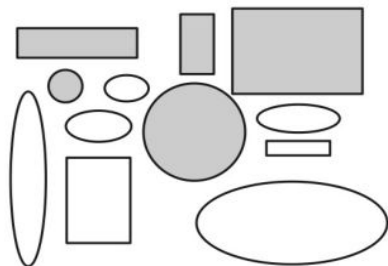


MDS ←

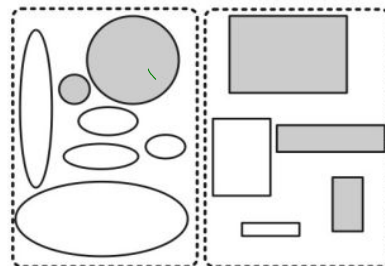
PCA

T-sne

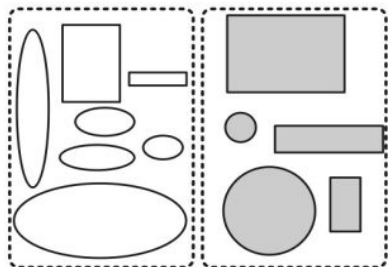
# Agrupamento de dados - subespaços



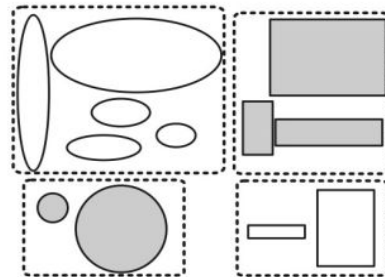
(a) Objetos



(b) Agrupamento pela forma (2 *clusters*)



(c) Agrupamento pelo preenchimento (2 *clusters*)



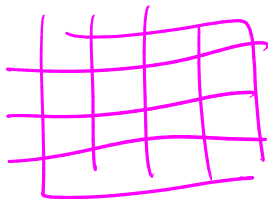
(d) Agrupamento pelo preenchimento e pela forma (4 *clusters*)

# Agrupamento de dados - paradigmas

Graças às definições pouco precisas e distintas, os variados critérios e a dificuldade em definir similaridade, há vários algoritmos de agrupamento, em diversos paradigmas.

# Agrupamento de dados - paradigmas

- Particionais *← k-means*
- Hierárquicos *←*
  - Aglomerativos ou divisivos *linkages*
- Baseados em densidade *DBSCAN, density-peaks*
- Baseados em grid *←*



# Agrupamento de dados - paradigmas

## Particionais

- Cria  $k$  ( $\ll n$ ) **grupos separados no espaço**, em que cada grupo possui ao menos um exemplo e **cada exemplo pertence a exatamente um grupo**

Exs:  $k$ -means,  $k$ -medoids, GMM



# Agrupamento de dados - paradigmas

## Hierárquicos

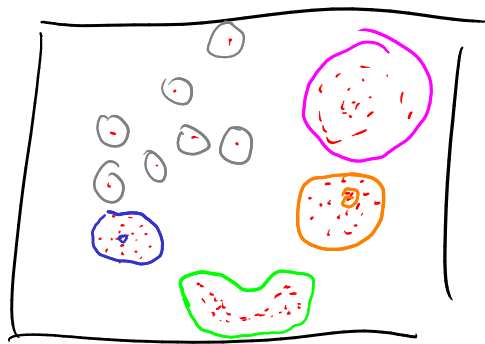
- Constrói uma **estrutura de hierarquia entre *clusters***, em que, a cada nível, um número diferentes de grupos é definidos. A hierarquia pode se dar pela junção de grupos (aglomerativos) ou pela divisão de grupos (divisivos).

Exs: single-, complete-, e average- linkages,  
*bisecting k-means*

# Agrupamento de dados - paradigmas

## Baseados em densidade

- A partir de **sementes** presentes em **regiões de alta densidade**, agrupa exemplos na vizinhança, até que se alcance regiões no espaço com baixa densidade, ou seja, que em um determinado raio a partir daquele exemplo, não haja mais que um determinado número de outros exemplos.

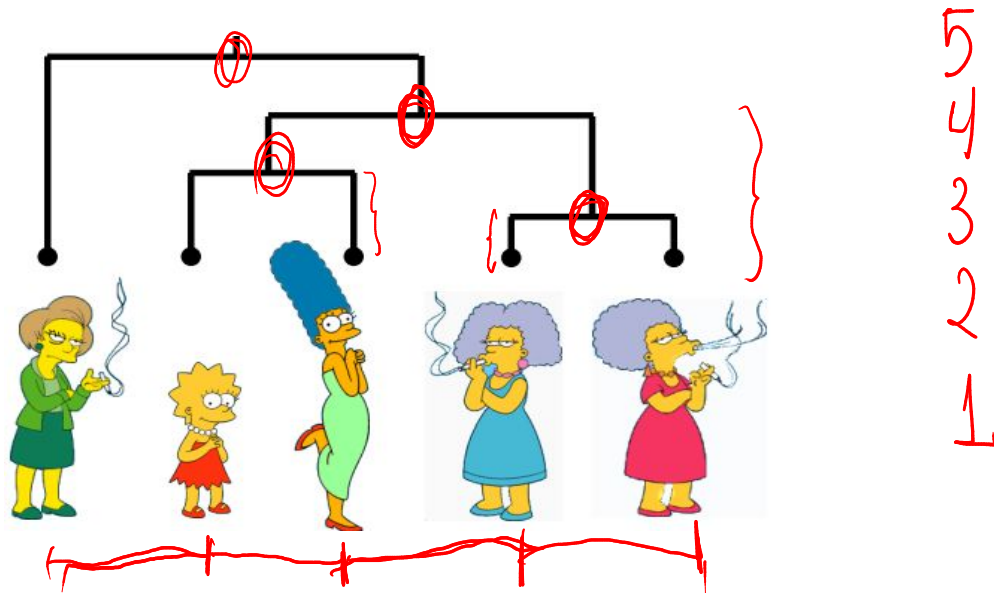


Exs: DBSCAN, *density peaks*

# Agrupamento hierárquico

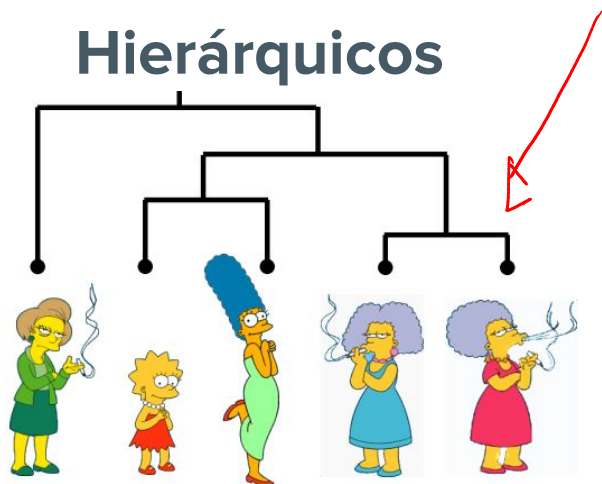
# Agrupamento hierárquico - Dendrograma

Forma mais comum de se representar uma hierarquia (de grupos)



# Agrupamento hierárquico - Dendrograma

Forma mais comum de se representar uma hierarquia (de grupos)



# Agrupamento hierárquico

Uma hierarquia é uma **sequência de partições aninhadas**

- Uma partição  $P_1$  está aninhada com uma partição  $P_2$  se cada grupo de  $P_1$  é um subconjunto de um grupo de  $P_2$

# Agrupamento hierárquico

Uma hierarquia é uma **sequência de partições aninhadas**

Ex:

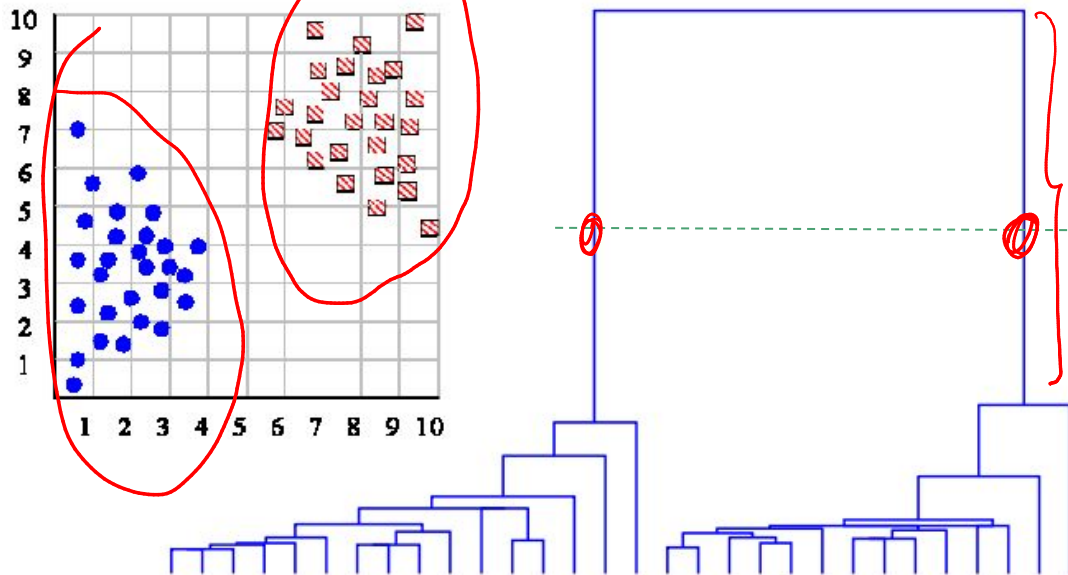
$$P1 = \{(x_1), (x_3, x_4, x_6), (x_2, x_5)\}$$

$$P2 = \{(x_1, x_3, x_4, x_6), (x_2, x_5)\}$$

$$P3 = \{(x_2, x_1), (\dots)\}$$

# Agrupamento hierárquico - Dendrograma

~~Pode auxiliar~~ a encontrar um número adequado de grupos

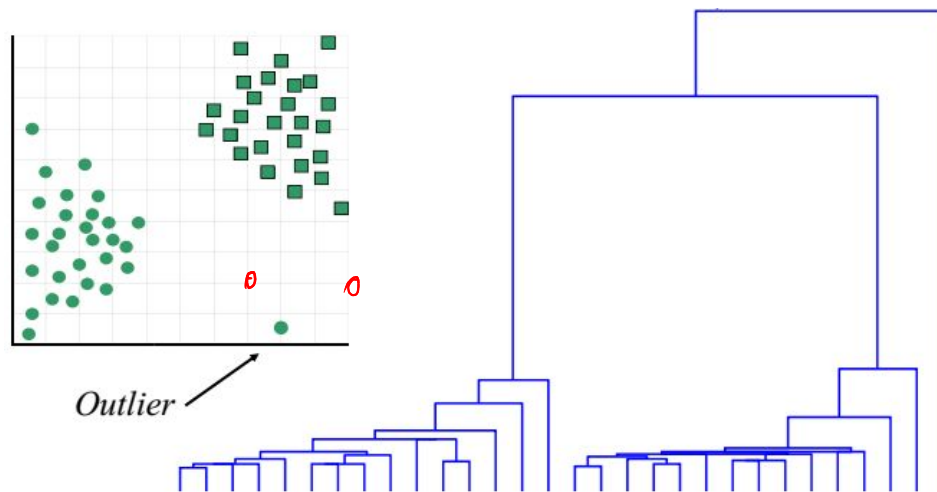




# Agrupamento hierárquico - Dendrograma

Pode auxiliar a encontrar *outliers* (ramo isolado)

- Nem sempre é fácil assim, na verdade



# Agrupamento hierárquico - Matriz de distâncias

Vários algoritmos de agrupamento hierárquico podem operar diretamente sobre a matriz de distâncias

- São, portanto, relacionais
- Diferentemente do *k-means*, mas similar ao *k-medoids*

Petitjean → averaging DTW

# Agrupamento hierárquico - Categorias

Há duas categorias de algoritmos:

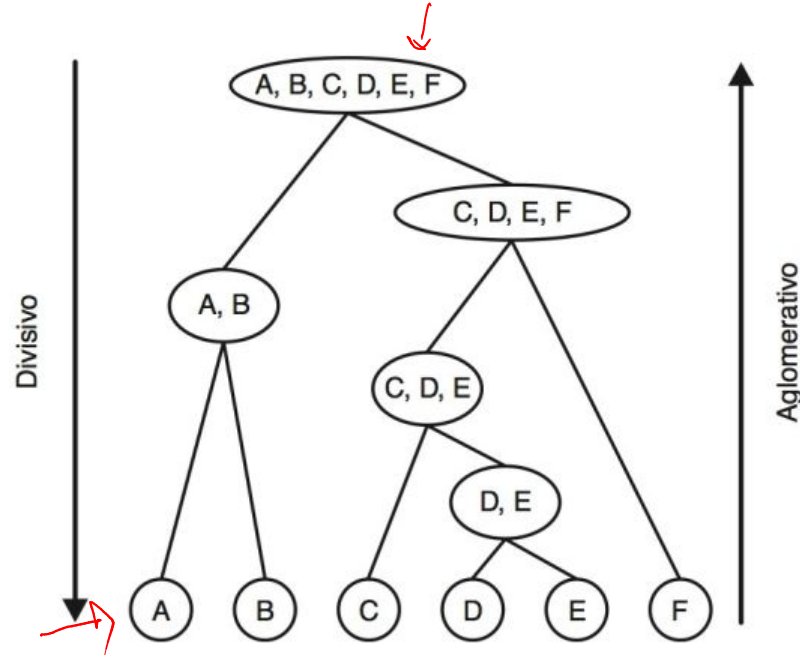
- Bottom-up (**aglomerativos**)
  - Se iniciam na “solução trivial”, ou seja, *singletons* ↵
  - Aglomeram/unem soluções parciais uma a uma, diminuindo em 1 o número de grupos a cada iteração

# Agrupamento hierárquico - Categorias

Há duas categorias de algoritmos:

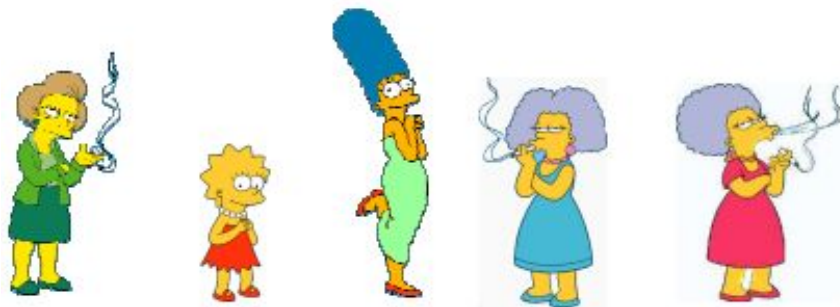
- Top-down (**divisivos**)
  - Se iniciam como um grupo com todos os exemplos
  - Divide a solução a cada iteração, gerando 2 novos grupos a partir de 1

# Agrupamento hierárquico - Categorias



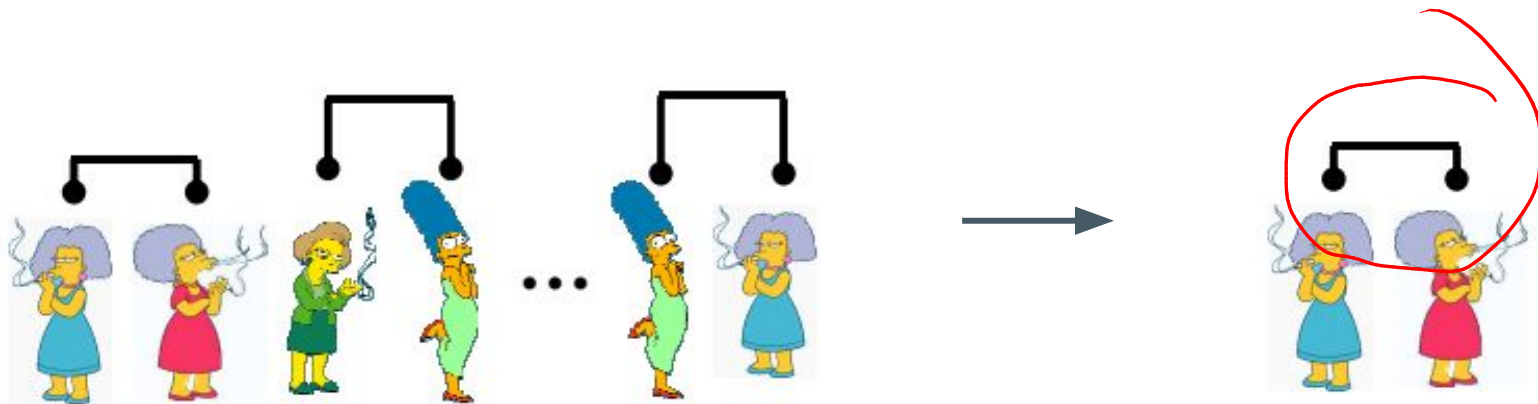
# Agrupamento hierárquico - Aglomerativos

Passo 1: considere cada exemplo como um cluster unitário (*singleton*), portanto há N conjuntos disjuntos



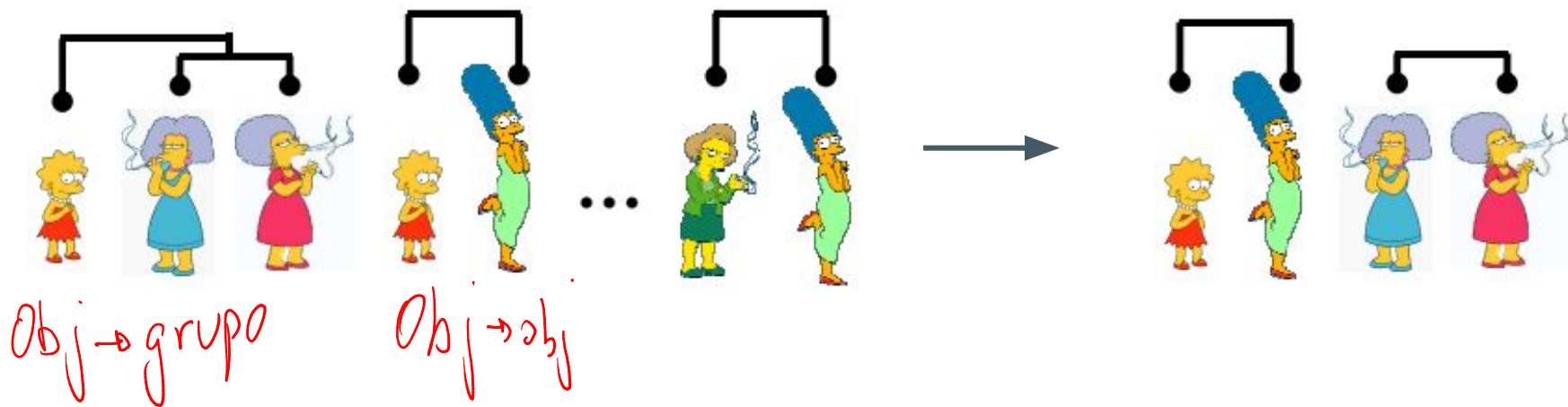
# Agrupamento hierárquico - Aglomerativos

Passo 2: escolha, de todas as ligações possíveis, a melhor segundo algum critério



# Agrupamento hierárquico - Aglomerativos

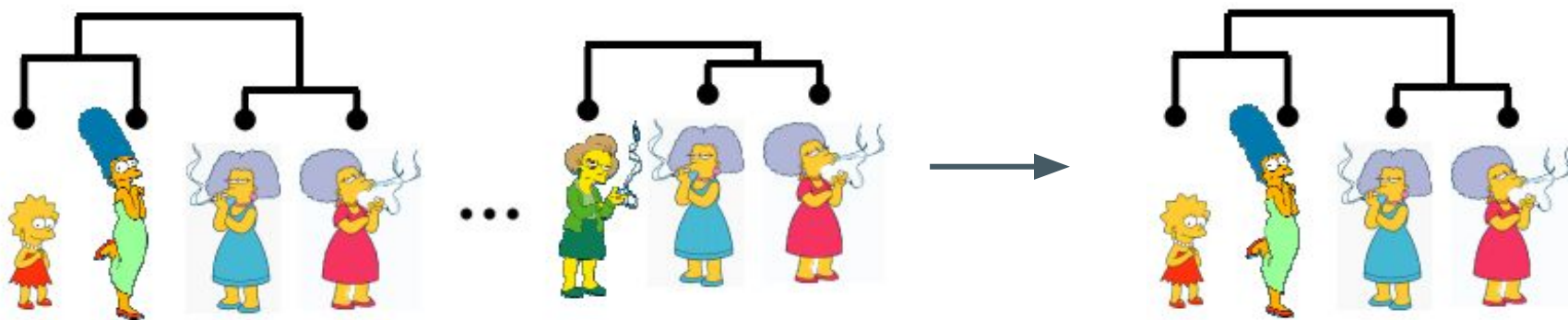
Passo 3: repita o passo 2 até ter apenas um conjunto





# Agrupamento hierárquico - Aglomerativos

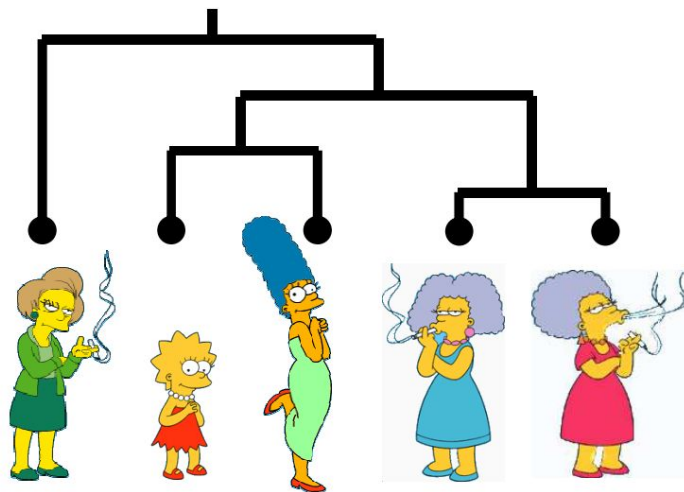
Passo 3: repita o passo 2 até ter apenas um conjunto



*grupo → grupo*

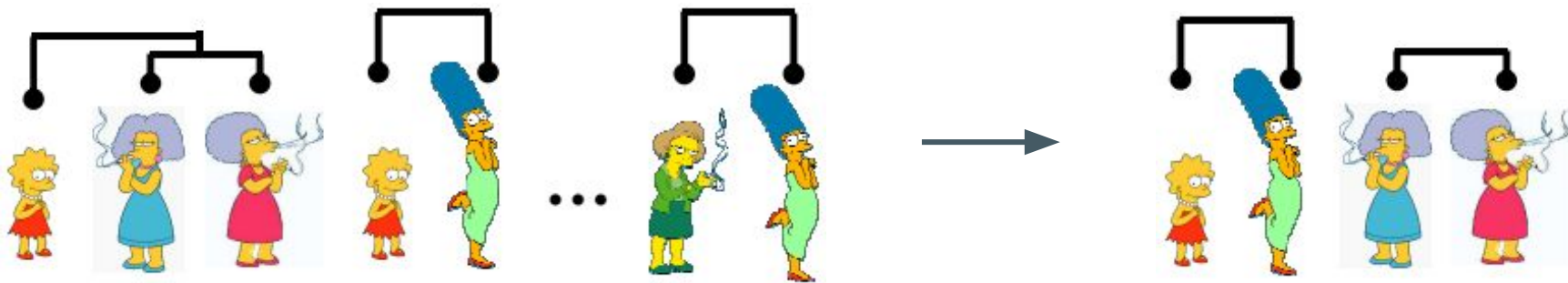
# Agrupamento hierárquico - Aglomerativos

Passo 3: repita o passo 2 até ter apenas um conjunto



# Agrupamento hierárquico - Aglomerativos

Perguntas: sabemos calcular a distância entre exemplos, mas... como faz para calcular a distância entre grupos? E entre grupos e exemplos?



# Agrupamento hierárquico - Aglomerativos

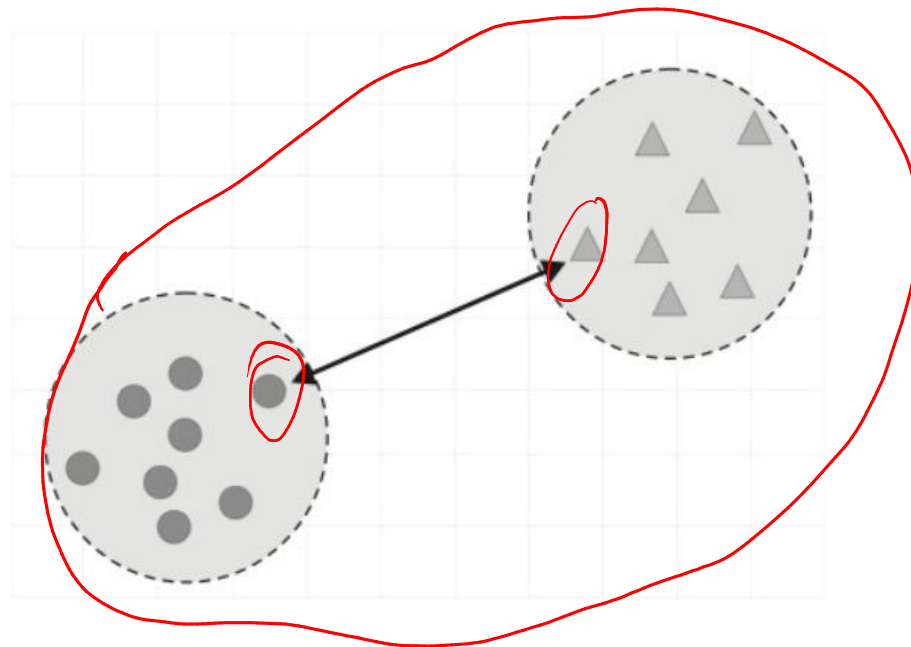
A resposta da pergunta anterior é o que define o algoritmo de agrupamento hierárquico aglomerativo.

- Single-linkage ✓
- Complete-linkage ✓
- Average-linkage ✓
- Método do centróide ✓
- Ward ✓
- ... ✓

# Agrupamento hierárquico - Single-linkage → MST } AGM }

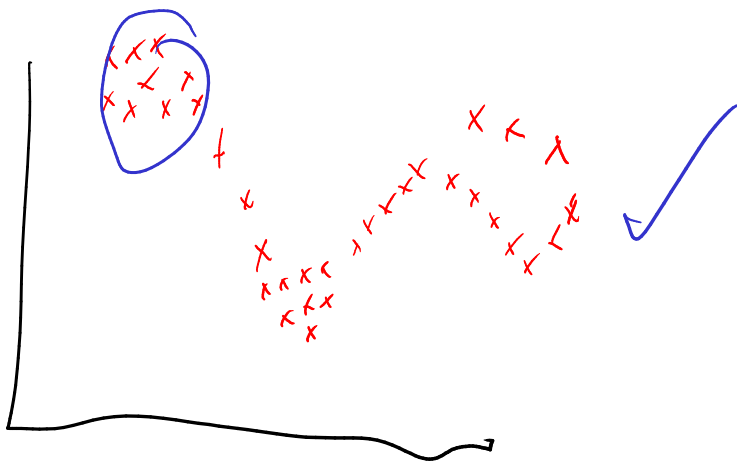
- A distância entre dois grupos é a menor distância entre dois pontos (um em cada grupo)
- Um exemplo é um grupo (unitário)
- A decisão de união é dada pela menor distância entre todos os pares de grupos

# Agrupamento hierárquico - Single-linkage

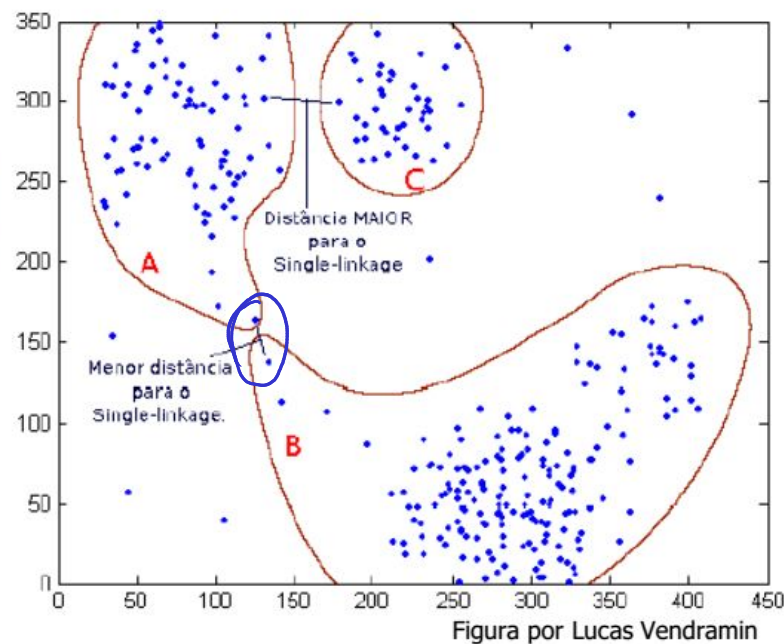
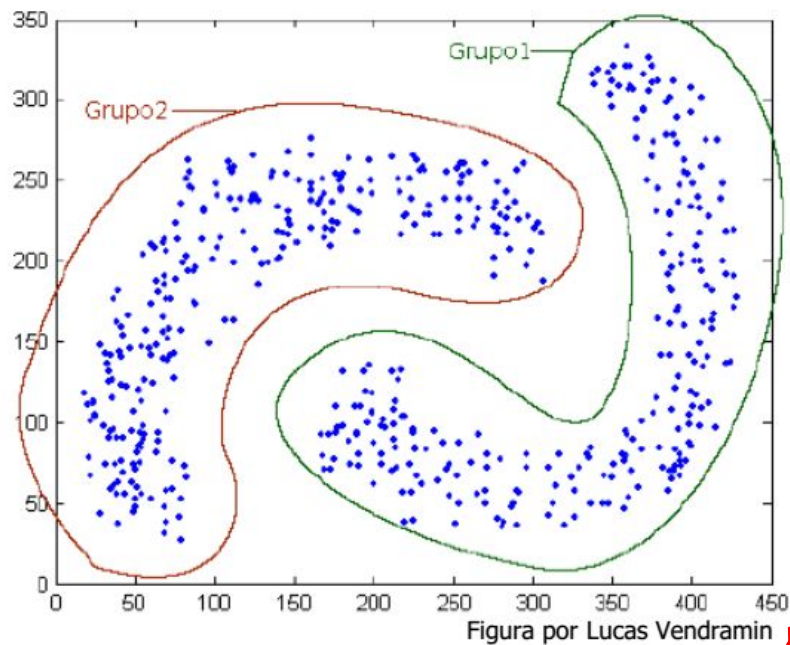


# Agrupamento hierárquico - Single-linkage

- Encontra formas arbitrárias
- Porém, é muito sensível a outliers e ruídos (ex: "pontes")



# Agrupamento hierárquico - Single-linkage





# Agrupamento hierárquico - Complete-linkage

- A distância entre dois grupos é a **maior distância entre dois pontos** (um em cada grupo)
- Um exemplo é um grupo (unitário)
- A decisão de **união é dada pela menor distância** entre todos os pares de grupos

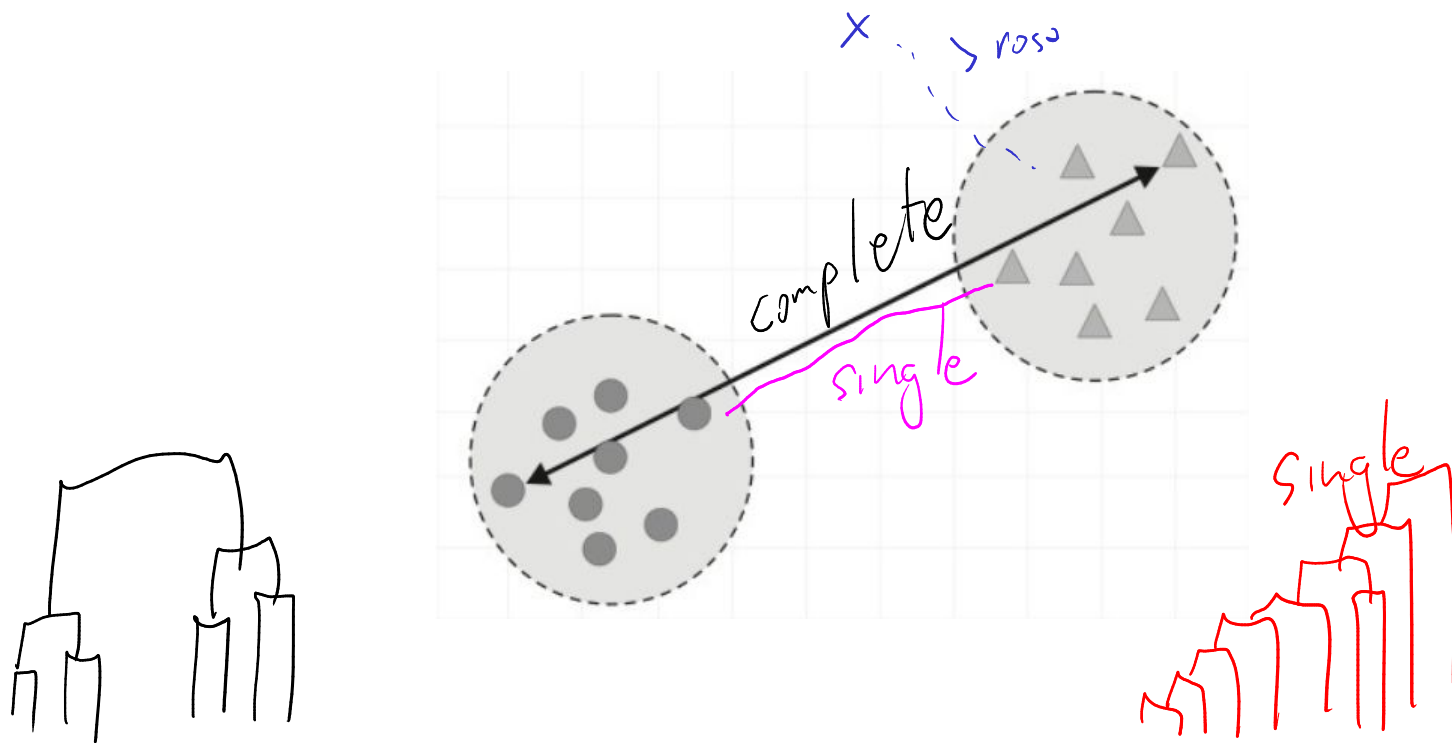
Também conhecido como **vizinho mais distante** ou *furthest neighbor*.

# Agrupamento hierárquico - Complete-linkage

Observação:

Este (e os próximos) tendem a encontrar grupos  
**hiper-esféricos.**

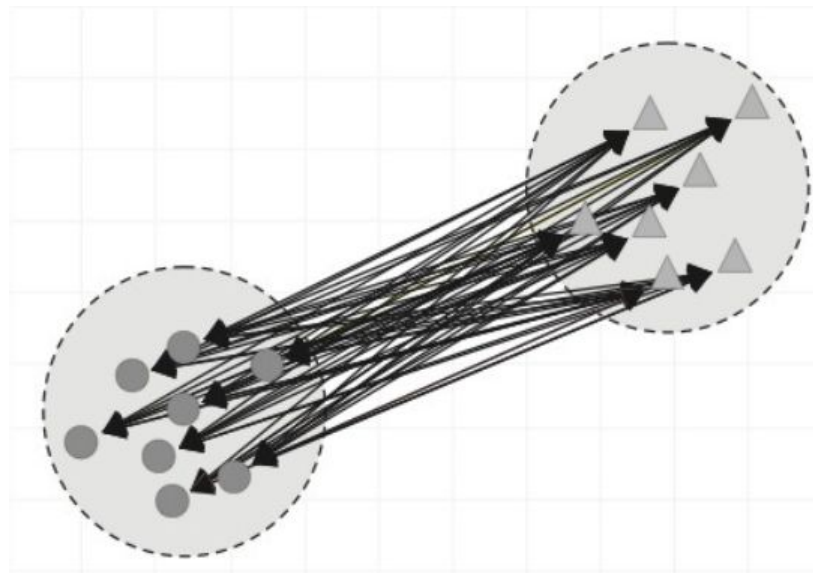
# Agrupamento hierárquico - Complete-linkage



# Agrupamento hierárquico - Average-linkage

- A distância entre dois grupos é a **distância média** entre os pares de pontos (nos grupos diferentes)
- Um exemplo é um grupo (unitário)
- A decisão de união é dada pela menor distância entre todos os pares de grupos

# Agrupamento hierárquico - Average-linkage



# Agrupamento hierárquico - Ward

- A distância entre dois grupos é dada pela variação no **critério J** da partição corrente se esses grupos forem unidos para formar a partição seguinte na sucessão hierárquica

$$J = \sum_{i=1}^k \sum_{x_j \in C_i} (x_j, \overline{x_i})$$

# Agrupamento hierárquico - Ward

- Unir os 2 grupos mais similares significa minimizar o crescimento das variâncias intra-grupos a cada nível da hierarquia

# Agrupamento hierárquico - Ward

Também encontra grupos hiper-esféricos, mas costuma ser mais robusto que os algoritmos apresentados anteriormente

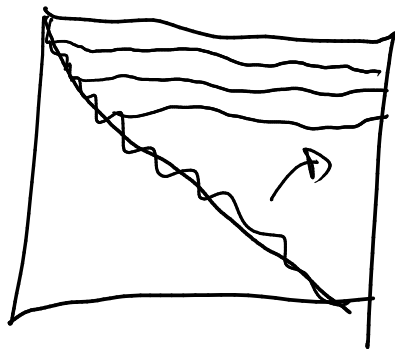


# Agrupamento hierárquico - Sumário

- Não é necessário definir número de grupos a priori ✓
- Problema de escalabilidade (confuso)
- Fácil interpretação (intuitiva) ✓
- Sensibilidade(s)

# Agrupamento hierárquico - Sumário

- Não é necessário definir número de grupos a priori
- Problema de escalabilidade ( $O(n^2)$ )
- Fácil interpretação (intuitiva)
- Sensibilidade(s)



$$d(x, x) = 0$$
$$d(x, y) = d(y, x)$$

# Critérios de validade relativos - Silhueta

**SWC:** silhueta média sobre todos os objetos

$$SWC = \frac{1}{N} \sum_{i=1}^N s(i)$$

$$s(i) = \frac{b(i) - a(i)}{\max[a(i), b(i)]}$$

$a(i)$  = distância média do  $i$ -ésimo objeto ao seu cluster

$b(i)$  = distância média do  $i$ -ésimo objeto ao cluster mais próximo

$s(i) = 0$  para *singletons*

$$SWC \in [-1, 1]$$